Virtual Institute of Microbial Stress and Survival
DOE Genomes To Life Project
Progress Report: November, 2003

I. Overview

The objective of this monthly progress report is to provide an update of the technical and administrative actions from the previous month as well as forecast upcoming progress for the VIMSS Genomes to Life Project at Lawrence Berkeley National Laboratory. I want to remind everyone how important to make sure everyone is communicating. The discussion boards (http://genomics.lbl.gov/~aparkin/discus) provide a forum for people to ask questions about direction of the project, priorities, and technical issues that can be read and answered by the entire group. I know email is often the most efficient means but it does privatize some of the important communications. Also, posting project data and information to BioFiles (https://tayma.lbl.gov/perl/biofiles) is EXTREMELY important. We are in the process of adding user help files to BioFiles – if you have user questions, please contact Keith Keller (tel: 510.495.2766 or email: kkeller@lbl.gov). This is the best metric I can give to the DOE leadership that we are making progress aside from the VIMSS website. Please make us and yourselves visible by donating data and information to the website.

II. Applied Environmental Microbiology Core

LBNL
SR-FTIR. We are setting up HPLC and GC/MS systems for exopolysaccaharide (EPS) analysis. This is important for the project because this would help to establish the SR-FTIR method, and to complement the metabolite analysis in Keasling's lab. Although EPS is a family of bacterial metabolites, which plays an important role in bacterial responses to stress factors, EPS is not currently being considered in Kealsing's effort. We are also preparing the manuscript "A Direct observation of *Desulfovibrio vulgaris* response to sudden influx of oxygen" for submission.

Biomass Production. Progress this month included the installation of two 4-liter FairMenTec bioreactors in our laboratory. Beto Zunigo from David Stahl's laboratory and Bill Yen from Judy Wall's laboratory visited mid month to install the units, due to the inability of the manufacturer of the bioreactors to send out an installation technician. Bill and Beto were invaluable and saved our labs weeks of frustration in trying to figure out the construction of the units. However, installation could not be completed due to the lack of a few key components. Thanks to some well-crafted emails from Nancy Slater, these are in route from Germany, and the installation should be completed mid – December. The new bioreactors will allow greater control of the conditions for cell growth. The goal is to run the bioreactors as chemostats, so that we have a continuous production of cells for the project.

Several filtration experiments using various types of glass fiber prefilters were conducted during the month of November. Although prefiltering helps increase the rate of filtration, we were not able to reach our desired goal of filtering 300 ml of culture in five minutes using the 47 mm filter holder. We decided to purchase a 142 mm filter holder because we believe that a larger filter holder with more surface area will result in less clogging and that prefiltering will not be necessary. We have also started researching the possibility of using a suitable tangential flow filtration (TFF) system as an alternative to centrifuging or normal flow filtration.

It would be helpful to receive feedback from groups if they could adapt their extraction procedures to accept cells collected on a large filter. The larger filters will allow for quicker sampling and also allow for tighter control on the sample (i.e. maintaining the control and stress conditions during sampling).

Continued development of growth curve measurements using 96 well plates is showing promise. One concern is the use of optical density to measure growth, especially when varying a medium component (e.g. Zn) where increased concentration may cause increase in precipitation. This month the Bradford protein assay was utilized at the end of the assay to confirm that the increase in optical density was correlated with increased biomass.

University of Washington

The stability of established methanogen-SRBs co-cultures (*Desulfovibrio vulgaris or Desulfovibrio sp*. PT2 with *M. maripulidisl)* was confirmed by serial transfer (five times) in McCm medium. The affect of $MgCl_2$ on co-culture growth was tested by modifying the $MgCl_2$ concentration to conform with that of the B3 medium. The greatest biomass and methane production for the *D. vulgaris* -*M. maripulidis* co-culture was observed in the B3 medium supplemented with magnesium chloride.   Similar biomass and methane yields were observed for the *Desulfovibrio sp*. PT2 -*M. maripulidis* co-culture in McCm and the amended B3 medium. Both co-cultures demonstrated slower growth and lower methane production in B3 without magnesium chloride (fig.1 a-c). Analyses of lactate and acetate concentrations at different time points in this experiment are in progress.

The affect of alternative reducing agents on the growth of *D. vulgaris* was investigated. *D. vulgaris* grows better in a medium amended with titanium chloride as the reducing agent than in media containing sulfide or cysteine (fig. 2).

We previously reported that we isolated colonies (sample FWB203-03d 04 from FRC area 2) in agar amended with lactate, acetate or a hydrogen and carbon dioxide. These were transferred to a liquid medium amended with the same corresponding electron donors and incubated at 30°Cand 25°C for three weeks. No growth has yet been observed in any of these cultures. We plan to evaluate growth using other media amended with different compounds such as amino acids and yeast extract.

To decrease the amount of air intrusion into the fixed bed biofilm reactor and the FairMenTec bioreactor, the gas distribution systems were modified by replacing tubing with copper pipes.

The reactor systems are now ready for further experimental evaluation. The fixed-bed bioreactor was modified to incorporate in-line "daughter" reactors for characterizing the kinetics of *Desulfovibrio* biofilm development.

Beto (U of W) and Bill Yen (U of Missouri) traveled to LBNL and worked with Sharon Borglin to install and discuss the FairMenTec bioreactors.

Figure 1. Growth and gas accumulation in co-cultures of SRB and *M. maripaludis*, 171hr of growth, 4th transfer
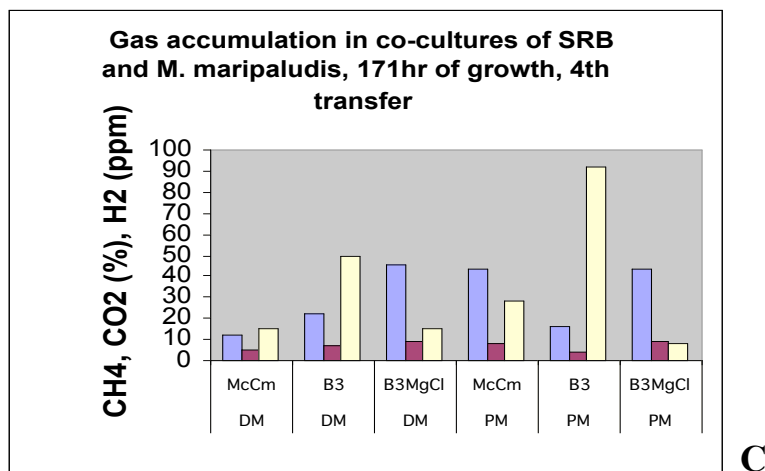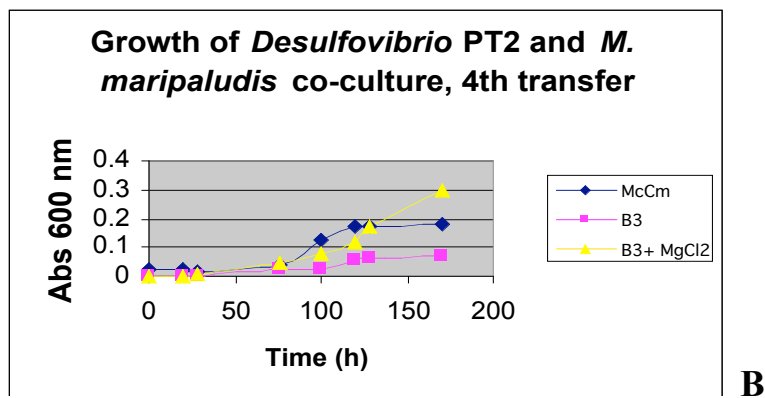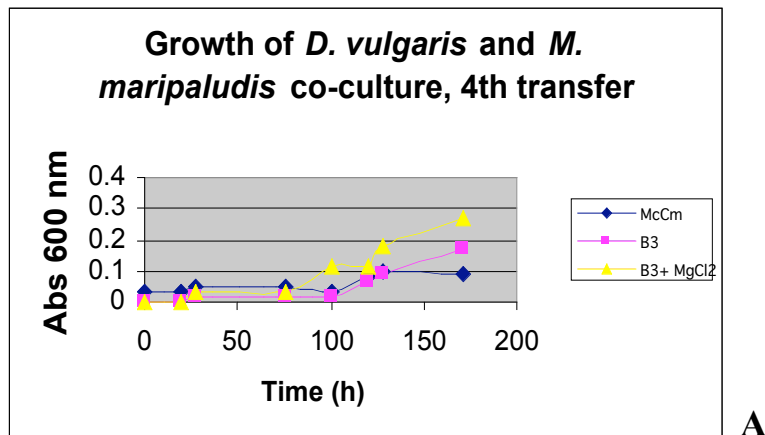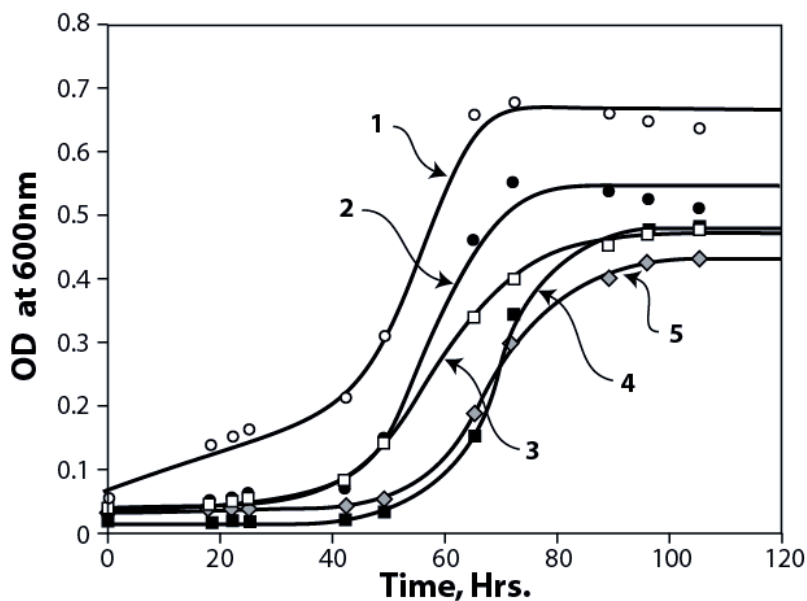
A

B

C

Figure 2. Effect of different reducing compounds on growth of *D. vulgaris*

1. TiCl$_3$, 0.058 g/
2. TiCl$_3$, 0.35 g/L
3. Na$_2$S, 2 mM
4. Cystein, 2 mM
5. Na$_2$S, 2mM and Cystein, 2mM

Oak Ridge National Laboratory

Draft manuscript of FRC sample analyses received from Matt Fields.

Diversa

Progress
The five samples described below have been amplified to obtain enough DNA for library construction. The amplified DNA has also been analyzed using diversity indexing to determine the complexity. Small insert library construction is complete for the five samples.

Genomes To Life Environmental Samples (October)

Sample ID	Large Fragment DNA Extraction	sample name	MDA	Diversity Indexing	DI Results	Small Insert Library	complete	Large Insert Library	complete
Area1 FB060-01-00	50g soil 10/1	1aL	10/8	10/17		10/10		10/10	
Area2 FB052-01-00	50g soil 10/1	2aL	10/8	10/17		10/10		10/10	
Area3 FB056-01-34	50g soil 10/1	3aL	10/8	10/17		10/10		10/10	
Area2 FB052-01-00	50g soil 10/2	2bL	10/8	10/17		10/10		10/10	
Area3 FB056-01-34	50g soil 10/2	3bL	10/8	10/17		10/10		10/10	

Issues

- Screening of libraries is dependent on selection of a suitable target.

Actions

- Large insert library construction is in progress on the above samples.

- Data is being analyzed from the sequencing QC of clones from the small insert libraries and from the diversity indexing.

III. Functional Genomics Core

Transcriptomics (ORNL)

Progress since last report

Shewanella
Raw microarray expression data for heat shock study were downloaded into Biofiles; computational core group at LBNL will send ORNL updated figures for the heat shock manuscript. Submitting the paper on salt shock stress will be a priority.

Desulfovibrio *whole-genome expression analysis*
The initial evaluations were performed using *D. vulgaris* cells grown at log and stationary phases of growth. Transcriptional analysis of *D. vulgaris* cells sampled during log phase growth indicated that 25% of annotated ORFs were up-regulated and 3% of annotated ORFs were down-regulated compared to stationary phase cells. The up-regulated genes included ORFs predicted to be involved with acyl chain biosynthesis, amino acid ABC transporter, translational initiation factors, and ribosomal proteins. In the stationary phase growth cells, the two most up-regulated genes (70-fold) were predicted to be a carboxynorspermidine decarboxylase and a 2C-methyl-D-erythritol-2,4-cyclodiphosphate (MECDP) synthase. Spermidines are polyamines that are typically abundant in rapidly dividing cells and are essential growth factors in eukaryotic organisms. Polyamines are thought to stabilize DNA by the association of the amino groups with the phosphate residues of DNA and can also enhance tRNA and ribosome stability. The MECDP synthase enzyme is essential in *Escherichia coli* and participates in the nonmevalonate pathway of isoprenoid biosynthesis, a critical pathway present in some bacteria and apicomplexans but distinct from that used by mammals. Several of the highly up-regulated ORFs were annotated as conserved hypothetical proteins. Interestingly, an ORF that was predicted to contain a flocculin repeat domain was almost 9-fold up-regulated in stationary phase growth cells compared to log phase growth cells. The flocculin domain has only been observed in fungi, and is thought to play a role during flocculation (non-sexual aggregation of single-cell microorganisms).

There were approximately 700 ORFs up-regulated and around 400 ORFs down-regulated when *D. vulgaris* cells were treated with 0.5 M NaCl for 0.5 hour. After 4-hour treatment, approximately 350 ORFs were seen to be up-regulated and more than 500 ORFs to be down-regulated. Patterns of gene expression were different. Interestingly, over half of the top 20 most up-regulated genes were conserved hypothetical or hypothetical genes.

Inhibition studies on *D. vulgaris* revealed an initial growth arrest after the addition of nitrite (5 or 10 mM). However, growth resumed gradually after 5 hours. Transcripts highly up-regulated throughout the 5 hours following nitrite shock included genes encoding two iron-sulfur cluster-binding proteins and a hybrid cluster protein. Notably, the nitrite reductase gene was moderately up-regulated (3 fold) only at 5 hours subsequent to nitrite shock.

Geobacter

- Construction of whole-genome *Geobacter* arrays are in progress. Some partial genome arrays have been printed and tested.

Proteomics (Diversa)

Objectives
- Proteomics analysis of *D. vulgaris* upon stress-response

Progress since last report
- Sample analysis
  - Repeating previous analyzed samples to obtain results with higher confidence
  - 3D LC MS/MS analysis are in progress (see below for details)

Digested Sample	Protein Extract (mg)	Digested Amount (ug)	Peptide samples	Sample name(2nd run)	Analysis time(2nd)
E3T0C1(X)	14 from 580	1000	Rg whole cell from D vulgaris E3T0C1	To be run	
E3T0C1(X)			Pellets after acid treated E3T0C1(X)	E3T0C1RgprHalf110403animal	2 days
E3T0V1(X)	15 from 440	1000	Rg whole cell from D vulgaris E3T0V1		
E3T0V1(X)			Pellets after acid treated E3T0V1(X)		
E3T1C1(X)	12 from 490	1000	Rg whole cell from D vulgaris E3T1C1	To be run	
E3T1C1(X)			Pellets after acid treated E3T1C1(X)	E3T1C1RgprHalf112403animal	2 days
E3T1V1(X)	7 from 580	1000	Rg whole cell from D vulgaris E3T1V1	To be run	
E3T1V1(X)			Pellets after acid treated E3T1V1(X)	E3T1V1RgprHalf120403animal	2 days

- Data analysis and data deposition
  - Finished all the database searching of the data from the first analysis (8 samples)
  - Completed preliminary data analysis
  - The results including the all the identified proteins and comparison are deposited on the GTL website.

Future work

- To complete 2nd 3D analysis
- To search all the data
- To combined the 1st and 2nd results
- To complete the data analysis

Protein complexes (Sandia)

Progress since last report

DIGE Optimization
Using MALDI PMF data obtained on our initial DIGE experiments (from July '03), we optimized parameters for database searching using Mascot. Parameters evaluated include spectrum processing (baseline correction, noise removal, deisotoping), peak intensity filtering, calibration type (internal vs external), and search mass window. Some parameters may need to be further optimized for "weak" data (i.e. spectra with poor S/N). Evaluated different MALDI sample preparation methods to maximize sensitivity for DIGE samples. Different recipes (various percentages of acetonitrile and/or methanol) and different commercial sources (Agilent and Sigma) for the CHCA matrix had little effect on sensitivity. However, the thin layer nitrocellulose spotting method (K. Perera, J. Perkins, and S. Kantartzoglou Rapid Communications in Mass Spectrometry 1995, 9, 180) improved the detection limit by 5-10x. We are currently using the optimized conditions for obtaining protein ids from the current set of DIGE runs.

LC-MS Optimization for Efficient Protein Complex Identification
Reproduced the experimental results reported by Russell et al (Anal. Chem. 2001, 73, 2682-2685.) for rapid proteolysis under organic solution conditions: Detectable PMFs for cytochrome c, ubiquitin, and myoglobin were observed after only 15-30 min proteolysis in 40% acetonitrile and after only 5 min in 80% acetonitrile. This has the potential application for protein proteolysis directly after LC separation and/or in rapid fashion for high throughput experiments involving protein complexes.

Abbreviations: MALDI = matrix assisted laser desorption ionization, PMF = peptide mass fingerprint, S/N = signal-to-noise, DIGE = differential in gel electrophoresis, CHCA = alpha-cyano-4-hydroxy-cinnamic acid, LC = liquid chromatography

Metabolomics and Proteomics (UCB, LBNL)

Progress since last report

Methods development for genetics in *D. vulgaris*.

We got the conjugation to work with pBMC7.

We figured out what substrate to use to see for β−galactosidase activity in the anaerobic chamber. It will be 4-Methylumbelliferyl β-D-galatoside (MUG). Stock solution of MUG was made in DMSO at 4mg/ml. 25 ul of this spread on LS4D plates was sufficient for S17-1(pBMC7) to show fluorescence in the chamber.

D. vulgaris may require the entire *lacZ* gene to show β-gal activity. pBMC7 has only lacZα. A broad host range vector was created by cloning the *lacZ* gene and oriT fragments into pBBR1 vector with the cat gene. Conjugation into *D.vulgaris* and β-gal assay remains to be done.

IV. Computational Core

LBNL-Arkin

In November, we finished a prototype of a "shopping cart" page to keep track of genes of interest to users of the VIMSS web tools. The prototype allows users to generate a report on genes they have added to their list, or to export gene sequences in FASTA format. In addition, users can save their lists and return to them at a later time by registering with the site and logging in. Over the next month, we hope to include options to perform multiple sequence alignments based on these gene lists, and to display phylogenetic trees based on those alignments. Within the next two months, we hope to add DNA motif-finding modules to the page, to provide a bioinformatics workbench so that biologists can use the VIMSS site for biological analysis as well as visualization of genomic data.

We have added a comparative genomic browser for the GO gene ontology to the VIMSS comparative genomics web tools. To our knowledge, this is the only tool that allows biologists to compare the genetic content of any number of genomes of their choice according to the well-defined GO hierarchy. To achieve this goal, we have implemented an automated pipeline to assign InterPRO domains and GO assignments to all of the genomes in the VIMSS DB. We expect these assignments to be finished by the end of December, at which point they will all be accessible from the main VIMSS site.

We have produced a preliminary workflow for the genome annotation jamboree in April, and are starting to incorporate draft sequences of the remaining sulfate-reducing bacteria into the VIMSS DB for automatic annotation. In November, Katherine Huang attended the TIGR genome annotation course to compare our proposed annotation plans with those being implemented at TIGR.

We have validated our operon predictions in 5 species using gene expression microarray data. Over the next month we will finish a draft of a paper describing the method and the test results. Among the improvements to the method this month, we have developed a novel procedure that improves predictions in genomes with a strong coding strand bias such as Bacillus subtilis.

Over the next month, we plan to implement major changes in the layout of the VIMSS comparative genomics website. These will include reformatting the protein pages to allow users to limit what information they want reported for each gene, and more comprehensive search page that allows users to search annotations, GO terms, and keywords associated with InterPRO domain assignments as well as gene symbols and common synonyms. By the end of the month, we plan to have a preliminary draft of three additional target genomes for April annotation in the DB including Desulfuromonas acetoxidans, Geobacter sulfureducens, and Desulfovibrio desulfuricans (G20).

LBNL-Dubchak

1. Analysis of regulator families

Three families of transcriptional regulators involved in the heavy metal resistance were considered (MerR, ArsR, 2CS)

The MerR family was analyzed in detail. It contains not only metal-dependent regulators, but also other types of regulators (e.g. SoxR). However, we demonstrated that the metal-dependent regulators of this family form a specific group and established rules for identification of new metal-dependent MerR-regulators.

Further, we analyzed MerR-family-dependent metal-resistance loci. In most cases (cadmium, copper, zinc, lead resistance), such loci contain a regulator (CadR, HmrR/CueR, ZntR, PbrR respectively) and a cation efflux transporter. However, the mercury-resistance loci are more complicated (MerR regulator and up to 8 enzymes, transporters, and conserved genes of unknown function), and their structure was described.

We also constructed formal rules for recognition of binding sites of metal-dependent regulators from the MerR family (separately for Gram-positive and Gram-negative bacteria, and for different regulators).

Thus, given a new genome, we believe we can completely characterize MerR-family-dependent systems of heavy-metal resistance.

Preliminary analysis of the ArsR family demonstrated that there are about 130 candidate metal-dependent regulators of this type. However, as the three genomes in question (See below) do not contain regulators of this family, this study was put on hold.

Some 2Cs (two-component system) metal-dependent systems were considered in detail. For some of them (copper resistance regulators PcoR, CusR, CopR from Pseudomonas and Ralstonia spp.) we constructed recognition rules for the candidate binding sites (so-caled COP-boxes).

2. Analysis of concrete genomes

Desulfovibrio desulfuricans

There are no metal-dependent MerR-family regulators, and no metal-type 2CS regulators. However, the genome contains four candidate cation efflux transporters, one of them likely to be copper exporter (the remaining three might be unrelated to heavy metals).

Geobacter metallireducens

There is one non-metal-dependent MerR-family transcriptional regulator, and no metal-dependent ones. The genome contains one candidate metal-dependent 2CS system (likely cobalt-zinc-cadmium one), but corresponding transporters have not been found in the neighborhood of these genes.

There are five cation transporters, one of which could be tentatively assigned the copper specificity.

Shewanella oneidensis

This genome has two cation transporters, one clearly a copper exporter, the other also might be a copper one. The former is a part of a divergon with a MerR-family regulator from the copper branch. A strong candidate binding site was observed in the intergenic region. Thus we have completely characterized the copper resistance locus of S. oneidensis: a MerR-family regulator gene and a transporter gene transcribed in opposite directions with the regulatory site in the intergenic region.

LBNL – Olken

Visual Graph Query User Interface

Viji examined a number of graph editor toolkits. After some discussion, we decided to adopt the use of JGraph as the toolkit used to implement the Visual Graph Query User Interface. Criteria for selection included code maturity and stability, existence of an ongoing development community, the ability to change icons, arrows, and use curved edges, nested graph capabilities, RDF and SVG export capability, multigraphs, etc. A number of alternative toolkits were considered: JGraph, Touchgraph, ZVTM library, GEF (Graph Editing Framework), ... We plan to add an integrated SVG viewing capability to JGraph, by including ZVGViewer from IsaViz. We will also need to add nested graph capability to JGraph.

We have constructed two menus using RSS Version 1.0 (an RDF based version).
We have not yet gotten the menus to pass RDF validation. We expect to complete basic menus in RSS in December and obtain some ontologies to encode in RDF.

Graph Data Model

Alex Gilman has begun work on a nested graph pathway editor and we will need nested graphs to encode complex queries. Therefore we have begun to look at various nested graph encodings in RDF. We hope to decide on a nested graph encoding in RDF this month (December).

We also plan to develop an initial RDF encoding template for Eric Alm's regulon dataset.

Navigational API

This API is being used in the browser development (see below). At this point, only packaging and documentation are needed for completion.

Constructing RDF Query Graphs

Construction of more elaborate RDF query graphs will require an RDF encoding of nested graphs (see above). Work on this has begun.

Browser

Development of the database schema and instance browser by Kevin Keck continued. Problems with using the browser with Internet Explorer appear to have been resolved. More documentation is still needed.

Preparing Test Sets for BGDM Initial Demo

This work involves converting B. subtilis dataset of Denise Wolf, et al. from the current DOT file format (used by Graphviz) to RDF.   Brian Carnes has converted these files to a simpler file format, which we will convert to RDF.

Graph Algorithm Development

Brian Carnes implemented a limited subgraph homeomorphism query algorithm in Prolog, which supports several of the queries that Denise Wolf is interested in. Kevin Keck has built a wrapper, which will take our RDF encoded queries and convert these to Carne's query language and then call Carne's code. The query conversion is complete, but conversion of the results to RDF is to be finished in December.
        We are continuing to examine alternative algorithms for subgraph homeomorphism.

Writing

The WDMBIO workshop report has been completed and copies were distributed at the NIH Digital Biology Meeting in the first week of November.   We still need to make arrangements to print and distribute additional copies and to construct a web page linking to the report and whitepapers.

Meetings, Collaborations

We (F. Olken) held a long discussion with Ron Taylor of PNNL concerning possible use of our biopathways database by PNNL. Taylor is definitely interested in using our software as soon as it is available. We anticipate that Taylor may be able to contribute useful criticism of our query language.

Plans for December 2003

1) Complete biopathways chapter for DOE Computational Biology Primer. (FO)
2) Code DOT file format (from ATT Graphviz) to RDF converter. (KK)
3) Specification of graph query language RDF encoding. (FO, KK)
4) Continue algorithm design for path queries which satisfy a regular expression. (FO)
5) Contact biopax.org about their efforts on standardization of biopathways data interchange format. (FO,KK)
6) Lecture on biopathways databases to BISC/FLINT meeting on Dec. 16 at 1:15 PM at Room 306, Soda Hall, UC Berkeley campus. (FO)
7) Complete wrapper for Brian Carnes subgraph homeomorphism query code. (KK)
8) Settle on RDF encoding for nested graphs (KK and FO).
9) Complete menu encodings in RSS/RDF. (FO)
10) Initial RDF encoding template of Eric Alm's regulon dataset. (FO)
11) Initial query drawing capability in VQGUI. (VN).

V. Project Management

Uploads to BioFiles: Each PI should designate a person in their laboratory to serve as the point of contact for the BioFiles data uploads. Data should be uploaded on a weekly basis or more frequently.

Updates to http://vimss.lbl.gov: There are several new feature on the VIMSS site including an updated calendar, job openings in which all of the GTL PIs can post, planned publications and there are plans for featuring a GTL scientist each month.

GTL Grantee Workshop: The DOE GTL Grantee Workshop will be held in Washington, DC on February 29-March 2. The Workshop website is http://www.orau.gov/gtl2004/. We will have four abstracts (one from each core group and one from Project Management) and one 4'x4' poster. All of the PIs are strongly encouraged to attend this Workshop.

VI. Planned Publications

1. Salt Stress paper by Jizhong Zhou at ORNL and LBNL Computational Core (January completion): Draft is complete – they are waiting for updated sequence information

2. pH Stress paper by Jizhong Zhou at ORNL and LBNL Computational Core (January completion): Draft is complete – they are waiting for updated sequence information

3. Temperature Stress paper by Jizhong Zhou at ORNL and LBNL Computational Core (January completion): Draft is complete – they are waiting for updated sequence information

4. Oxidative Stress paper by Jizhong Zhou at ORNL and LBNL Computational Core (January completion)

5. Biofilms verses planktonic cell growth paper by David Stahl: Terry and David are going to rough out an outline

6. Nature Insight paper by Adam Arkin (November completion)

7. Characterization of microbial communities under the stressful conditions of nitrate, heavy metals, and acidic pH by Mathew Fields (late November completion)

8. Review of metabolomics paper by Vince Martin (January 04 completion)

9. "A real-time observation of Desulfovibiro vulgaris response to oxygen shock" by Hoi-Ying Holman (Draft in November)

10. "X-ray tomorgraphy of Desulfovibiro vulgaris in oxygen stress" by Hoi-Ying and Carolyn Larabell (Spring 2004)

11. Stress in Metal-reducing Bacteria Review by Terry Hazen and David Stahl

12. Research paper by Lihong Sun (late December completion)

13. Proteomics, 2D-Gel and O_2 Stress paper by Anup and Diversa (late December completion)

14. Protease Heat Shock paper by Anup and Diversa (late December completion)

15. The origin and maintenance of operon structure (late December completion)

16. An unsupervised learning approach to operon identification (late December completion)

17. VIMSS DB: an open source database and applications suite for comparative genomics (late December completion)